Excursion and contour uncertainty regions for latent Gaussian models

DAVID BOLIN[1] AND FINN LINDGREN[2]

[1]*Centre for Mathematical Sciences, Lund University, Lund, Sweden*
[2]*Mathematical Sciences, University of Bath, Bath, United Kingdom*

Abstract: An interesting statistical problem is to find regions where some studied process exceeds a certain level. Estimating such regions so that the probability for exceeding the level in the entire set is equal to some predefined value is a difficult problem that occurs in several areas of applications ranging from brain imaging to astrophysics. In this work, a method for solving this problem, as well as the related problem of finding uncertainty regions for contour curves, for latent Gaussian models is proposed. The method is based on using a parametric family for the excursion sets in combination with a sequential importance sampling method for estimating joint probabilities. The accuracy of the method is investigated using simulated data and two environmental applications are presented. In the first application, areas where the air pollution in the Piemonte region in northern Italy exceeds the daily limit value, set by the European Union for human health protection, are estimated. In the second application, regions in the African Sahel that experienced an increase in vegetation after the drought period in the early 1980s are estimated.

Key words: Latent Gaussian models; excursion sets; contour curves; multiple testing

1 Introduction

In many statistical applications, one is interested in finding areas where the studied process exceeds a certain level or is significantly different from some reference level. A typical example is in studies of air pollution, where one is interested in testing if, and where, the pollution level exceeds some given limit value set by some regulatory agency [Cameletti et al., 2012], and similar examples can be found in a wide range of scientific fields including brain imaging [Marchini and Presanis, 2003] and astrophysics [Beaky et al., 1992]. In spatio-temporal applications one might be interested in finding regions that have experienced significant changes over the studied time period. This is a common problem in climate science and the studied quantity can for example be temperature [Furrer et al., 2007], precipitation [Sain et al., 2011], or vegetation [Eklundh and Olsson, 2003, Bolin et al., 2009].

The quintessential problem is that one has observations $\mathbf{y}$ of some latent stochastic field $x(\mathbf{s})$ and wants to find a region D such that, with a certain given probability, $x(\mathbf{s}) > u$ for all $\mathbf{s} \in D$ for a given level u. The easiest, and most common, way of specifying D is to choose it as the set of locations

$$D_m = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) \geq 1 - \alpha\}, \tag{1}$$

where the probability is taken under the posterior distribution for $x|\mathbf{y}$. Thus, D_m is specified as the set of locations where the marginal probability for exceeding the level exceeds some given value $1 - \alpha$. The set can be calculated using pointwise hypothesis testing and the parameter α then acts as the type 1 error parameter, and thus controls how confident one is that the level is exceeded at each location. The problem with this definition of D is that of multiple hypothesis testing; the confidence level α does not give us any information about the family-wise error rate, and hence does not quantify the certainty of the level being exceeded at all points in the set simultaneously. That is, the probability $\mathsf{P}(x(\mathbf{s}) > u, \mathbf{s} \in D_m)$ is in general smaller than $1 - \alpha$. If one instead wants to choose D so that this simultaneous probability is $1 - \alpha$, one has to modify the procedure for constructing the set.

The more general problem of multiple hypothesis testing is an active research area and there exists a number of proposed solutions for problems in various contexts. Most of these solutions are based on first calculating the marginal probabilities $\mathsf{P}(x(\mathbf{s}) > u)$, then calculating a single threshold value t, and finally defining the exceedance region as $D = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) > t\}$. The methods differ in

how the threshold t is calculated, and can basically be divided into three main categories; type 1 error control thresholding, false discovery rate thresholding, and posterior probability thresholding [Marchini and Presanis, 2003]. The most popular method is likely the method by Adler [1981] using the Euler characteristic of the latent field to control the family-wise error rate when defining the threshold t. Though this method is simple to use, one has to be careful to check whether the required assumptions are satisfied. Typically the method is accurate for large values of u, and the latent field is assumed to be stationary.

In this work we will focus on the problem where the latent spatial field $x(\mathbf{s})$ is Gaussian and measured at a set of irregular locations. This means that the posterior distribution $\pi(x|\mathbf{y})$ is non-stationary and typically non-Gaussian unless the measurements are Gaussian and the model parameters are known a priori. One of our motivating examples is the problem of finding regions with significant changes in vegetation in the African Sahel studied by Eklundh and Olsson [2003] and Bolin et al. [2009], and in this example the threshold u is zero, so methods based on asymptotic arguments when u goes to infinity are unlikely to perform well.

The method derived here is based on using a parametric family for the excursion sets in combination with a sequential importance sampling method for estimating joint probabilities. For a specific choice of the parametric family, the method is equivalent to the thresholding methods mentioned above, with the important difference that the correct joint distribution is used when selecting the threshold. The method is extended using more general parametric families, and the related problem of finding uncertainty regions for contour curves is treated using the same methodology.

The structure of the article is as follows. In Section 2, the problem is formulated and definitions for excursion sets and uncertainty regions for contour curves are given. In Section 3, a method for estimating these sets is proposed. Estimating the sets is the most difficult problem as one easily runs into computational difficulties arising from having to evaluate high-dimensional integrals. In Section 4, the methods are tested on a few simulated examples to test the method's accuracy. Two applications to real data are covered in Section 5, the first considers air pollution data from the North-Italian region Piemonte, and the second considers estimation of spatially dependent vegetation trends in the African Sahel. Finally, a few remarks and comments are given in Section 6.

2   Problem formulation

There are a number of different ways one could formulate excursion sets, and not all of them are useful from a practical point of view. Hence, in this section we will formalise the problem and discuss how the results should be interpreted. More precisely we look at two connected problems. The first one is to find areas where a stochastic process exceeds a given level with some probability and the second one is to quantify the uncertainty in contour curves of stochastic fields.

Throughout this section, let Ω be a bounded domain of $\mathbb{R}^n$, or have a well-defined area $|\Omega| < \infty$. First some notation for excursion sets of a function and contour sets is needed.

Definition 2.1 (Excursion sets for functions)**.** Given a function $f(\mathbf{s})$, $\mathbf{s} \in \Omega$, the positive excursion set A_u^+ for a level u is given by

$$A_u^+(f) = \{\mathbf{s} \in \Omega; f(\mathbf{s}) > u\}.$$

Similarly

$$A_u^-(f) = \{\mathbf{s} \in \Omega; f(\mathbf{s}) < u\}.$$

is the negative excursion set.

In a similar fashion one could now define the set of contour points for the level u as the set of points $\mathbf{s}$ for which $f(\mathbf{s}) = u$; however, a contour curve consists not only of these points but also

discontinuous crossings of the level u. In order to incorporate both continuous and discontinuous crossings, a contour point is defined as a point $\mathbf{s}$ such that in every neighborhood B of $\mathbf{s}$

$$\exists\, \mathbf{s}_1, \mathbf{s}_2 \in B : f(\mathbf{s}_1) \geq u, f(\mathbf{s}_2) \leq u.$$

The set of all such points is the complement of the union of the interior sets of the positive and negative excursion sets.

Definition 2.2 (Contour sets for functions). Given a function $f(\mathbf{s})$, $\mathbf{s} \in \Omega$, the contour set A_u^o for a level u is given by

$$A_u^c(f) = \left(A_u^+(f)^o \cup A_u^-(f)^o \right)^c.$$

where A^o is the interior, relative to Ω, of the set A and A^c is the complement.

Remark 1. Taking the interiors of the sets $A_u^+(f)$ and $A_u^-(f)$ is important. Consider for example the following function on $\Omega = [0, 1]$

$$f(s) = \begin{cases} -1 & 0 \leq s < 0.5 \\ 1 & 0.5 \leq s \leq 1. \end{cases}$$

In this case $A_0^+(f) \cup A_0^-(f) = \Omega$, so without taking the interiors of the sets $A_0^c(f)$ would be empty when we want to include the discontinuous crossing at 0.5 in the contour set. It is also important to take the interiors with respect to Ω and not $\mathbb{R}$, since the endpoints 0 and 1 always would be included in the contour set otherwise. This may seem as only a theoretical nicety, but the problem with discontinuous functions occurs frequently in environmental applications when discontinuous covariates are used for the mean value function of the field. This makes it essential to not treat contour sets as regions where the function lies close to a level, but rather as regions where level *crossings* occur.

The statistical problem is now to find a region D such that the function $x(\mathbf{s})$ exceeds the level u with a certain probability $1 - \alpha$ for all $\mathbf{s} \in D$. There might be many such regions, so if one is interested in a single answer one might look for the largest of these.

Definition 2.3 (Excursion sets). Let $x(\mathbf{s})$, $\mathbf{s} \in \Omega$ be a random field (or process). The positive level u excursion set with probability $1 - \alpha$ is given by

$$E_{u,\alpha}^+(x) = \arg\max_D \{|D| : \mathsf{P}(D \subseteq A_u^+(x)) \geq 1 - \alpha\}.$$

Similarly

$$E_{u,\alpha}^-(x) = \arg\max_D \{|D| : \mathsf{P}(D \subseteq A_u^-(x)) \geq 1 - \alpha\}.$$

is the negative level u excursion set with probability $1 - \alpha$.

Remark 2. The set $E_{u,\alpha}^+(x)$ can also be formulated as the largest set D for which $\mathsf{P}(\inf_{\mathbf{s} \in D} x(\mathbf{s}) \leq u) \leq \alpha$, which can be useful when calculating the set in practice. Also note that for deterministic functions f one has $E_{u,\alpha}^+(f) = A_u^+(f)$ and $E_{u,\alpha}^-(f) = A_u^-(f)$ for any $\alpha \in [0, 1]$.

It is important to realize how the excursion set $E_{u,\alpha}^+(x)$ should be interpreted: It is the largest set so that the level u is exceeded *at all locations* in the set with probability $1 - \alpha$, and therefore is a smaller set than D_m defined in (1), which is the set of points where the marginal probability for exceeding the level is at least $1 - \alpha$. Another possible definition of an excursion set would be a set that contains *all excursions* with probability $1 - \alpha$. This is a larger set than D_m, given by $E_{u,\alpha}^-(x)^c$. Which set one is interested in depends on the application, but it can be a good idea to calculate both to get a better understanding of the uncertainties in the problem.

In certain applications, one might be interested in joint positive and negative excursions from some level, for example when doing simultaneous regressions and one is interested in finding regions where the slopes are significantly different from zero (see Section 5.2 for a possible scenario of this kind).

Definition 2.4 (Level avoiding sets). Let $x(\mathbf{s})$, $\mathbf{s} \in \Omega$ be a random field. The pair of level u avoiding sets with probability $1 - \alpha$ is given by

$$(M_{u,\alpha}^+(x), M_{u,\alpha}^-(x)) = \underset{(D^+, D^-)}{\arg\max}\{|D^- \cup D^+| : \mathsf{P}(D^- \subseteq A_u^-(x),\, D^+ \subseteq A_u^+(x)) \geq 1 - \alpha\}.$$

Denote the union of these two sets the level avoiding set $M_{u,\alpha}$:

$$M_{u,\alpha}(x) = M_{u,\alpha}^+(x) \cup M_{u,\alpha}^-(x).$$

Remark 3. The sets $M_{u,\alpha}^+(x)$ and $M_{u,\alpha}^-(x)$ must be non-overlapping for the probability to be non-zero. The set $M_{u,\alpha}$ can be calculated as an excursion set itself. To see this, define a new random process $y(\mathbf{s})$ by

$$y(\mathbf{s}) = \begin{cases} u - x(\mathbf{s}), & \mathbf{s} \in D^-, \\ x(\mathbf{s}) - u, & \mathbf{s} \notin D^-. \end{cases}$$

The probability calculation in Definition 2.4 can now be reformulated as an ordinary excursion probability in y:

$$\mathsf{P}(D^- \subseteq A_u^-(x),\, D^+ \subseteq A_u^+(x)) = \mathsf{P}(D^- \cup D^+ \subseteq A_0^+(y)).$$

Also in this case, the set can be found using a reformulation using the infimum over the region as $\mathsf{P}(\inf_{\mathbf{s} \in D^- \cup D^+} y(\mathbf{s}) \leq 0) \leq \alpha$.

Similarly to how the contour sets for deterministic functions were defined, the pair of level avoiding sets can now be used to define uncertainty regions for contour curves.

Definition 2.5 (Uncertainty region for contour sets). Let $x(\mathbf{s})$, $\mathbf{s} \in \Omega$ be a random field, and let $(M_{u,\alpha}^+(x), M_{u,\alpha}^-(x))$ be the pair of level avoiding sets from Definition 2.4. The set

$$M_{u,\alpha}^c(x) = \left(M_{u,\alpha}^+(x)^o \cup M_{u,\alpha}^-(x)^o\right)^c$$

is then an uncertainty region for the contour set of level u.

The interpretation of this uncertainty region is important. The set $M_{u,\alpha}^c$ is the smallest set such that *all* level u crossings of x are in the set with probability $1 - \alpha$. One should note that this definition of the uncertainty region for level curves is different from some other definitions in the literature. For example, Lindgren and Rychlik [1995] define uncertainty regions as a union of intervals where each interval contains a single level crossing with probability $1 - \alpha$.

It is somewhat unsatisfactory that the sets defined here are made unique by finding the largest set satisfying a certain restriction. The set $E_{u,\alpha}^+(x)$ is for example defined as the largest set D satisfying $\mathsf{P}(D \subseteq A_u^+(x)) \geq 1 - \alpha$, but there are also many other smaller sets satisfying the requirement, and these are not seen if only $E_{u,\alpha}^+(x)$ is reported. Also, if one wants to know where the field likely exceeds the level u, the set $E_{u,\alpha}^+(x)$ might not be sufficient since it does not provide any information about the locations not contained in the set. Therefore, it would be good to have something similar to p-values, i.e. the marginal probabilities of exceeding the level, but which can be interpreted simultaneously. To that end we introduce the excursion function, level avoidance function, and contour function as visual tools for answering such questions.

Definition 2.6 (Excursion functions). The positive and negative u excursion functions are given by

$$F_u^+(\mathbf{s}) = \sup\{1 - \alpha; \mathbf{s} \in E_{u,\alpha}^+\},$$
$$F_u^-(\mathbf{s}) = \sup\{1 - \alpha; \mathbf{s} \in E_{u,\alpha}^-\}.$$

Similarly, the level avoidance and contour functions are given by

$$F_u(\mathbf{s}) = \sup\{1 - \alpha; \mathbf{s} \in M_{u,\alpha}\},$$
$$F_u^c(\mathbf{s}) = 1 - F_u(\mathbf{s}).$$

These functions take values between zero and one, and for a fixed level u and a fixed location $\mathbf{s}$, this value is equal to $1 - \alpha$ for the smallest α such that the location is a member of the excursion set. Thus, the set $E_{u,\alpha}^{\bullet}$ can be retrieved as the $1 - \alpha$ excursion set of the function $F_u^{\bullet}(s)$. The interpretation of the excursion function is therefore that if, for a given location $\mathbf{s}$, the function takes a value close to one, this indicates that this location is a member of the excursion sets for almost all values of α, whereas if the value of the function is close to zero, the location is only a member of excursion sets with large values of α and it is therefore more unlikely that the process exceeds the level at that location.

3 Computations

So far, no assumptions have been made regarding the distribution of $x(\mathbf{s})$, but to be able to calculate the excursion sets in practice we will now restrict ourselves to the class of latent Gaussian models, which is a popular model class with many practical applications [see e.g. Rue et al., 2009]. Thus, the following problem setup is assumed. Let $x(\mathbf{s})$ be a random field that can be written on the form

$$x(\mathbf{s}) = \sum_{i=1}^{k} \beta_i f_i(\mathbf{s}) + z(\mathbf{s})$$

where $f_i(\mathbf{s})$ are fixed effects and $z(\mathbf{s})$ is a zero mean random field with covariance parameters $\boldsymbol{\theta}_1$. Further assume that both $z(\mathbf{s})$ and the parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^{\top}$ are a priori Gaussian.

Let $\mathbf{y} = (y_1, \ldots, y_m)^{\top}$ be a vector of measurements of the latent field with some distribution $\pi(\mathbf{y}|\mathbf{x}_{obs}, \boldsymbol{\theta}_2)$, where $\mathbf{x}_{obs}$ is a vector containing the latent field evaluated at the measurement locations and $\boldsymbol{\theta}_2$ is a vector of parameters for the measurement distribution. Finally let $\mathbf{s}_1, \ldots, \mathbf{s}_n$ be the set of locations where predictions of the latent field should be calculated and let $\mathbf{x} = (x(\mathbf{s}_1), \ldots, x(\mathbf{s}_n))^{\top}$. The posterior distribution for $\mathbf{x}$ can then be written as

$$\pi(\mathbf{x}|\mathbf{y}) = \int \pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}) \, \mathrm{d}\boldsymbol{\theta}, \tag{2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\top}, \boldsymbol{\theta}_2^{\top})^{\top}$, and this is the distribution that should be used in the probability calculations when estimating the excursion sets.

There are now, in principle, two main problems that have to be solved in order to find the excursion sets, level avoidance sets, or contour uncertainty sets:

integration For excursion sets, calculate $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$ or $\mathsf{P}(D \subseteq A_u^-(\mathbf{x}))$ for a given set D, or in the case of level avoidance sets or uncertainty regions for contour curves, calculate $\mathsf{P}(D^- \subseteq A_u^-(\mathbf{x}), D^+ \subseteq A_u^+(\mathbf{x}))$ for the pair of sets (D^+, D^-).

optimization Use shape optimization to find largest region D satisfying the required probability constraint.

Hence, given a method to solve each of the two problems, one could simply run the shape optimization algorithm and in each iteration calculate the required probability using the integration method. In theory there are no problems doing this, but in practice the integration method will be computationally demanding and it may not be feasible to use this strategy for applications involving large data sets. Therefore, we instead propose a slightly different strategy that will minimize the number of calls to the integration method by solving the problem sequentially. We first outline the strategy in the simplest possible situation, which will be used as a basis for all other more complicated strategies.

The method is based on using an increasing parametric family for the excursion sets in combination with a sequential integration routine for calculating the probabilities. The advantage with using a sequential integration routine is that if the required probability has been calculated for some set D_1, then the calculation for a larger set $D_2 \supset D_1$ can be based on the result for D_1, resulting in large computational savings.

Algorithm 3.1 (Calculating excursion sets using a one-parameter family)**.** *Assume that the model parameters $\boldsymbol{\theta}$ are known and that the posterior distribution $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$ is Gaussian. Further assume that $D(\rho)$ is a parametric family for the possible excursion sets, such that $D(\rho_1) \subseteq D(\rho_2)$ if $\rho_1 < \rho_2$. The following strategy is then used to calculate $E_{u,\alpha}^+$.*

1. *Choose a suitable (sequential) integration method for the problem.*

2. *Reorder the nodes to the order they will be added to the excursion set when the parameter ρ is increased.*

3. *sequentially add nodes to the set D according to the ordering given above and in each step update the probability $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$. Stop as soon as this probability falls below $1 - \alpha$.*

4. *$E_{u,\alpha}^+$ is given by the last set D for which $\mathsf{P}(D \subseteq A_u^+(\mathbf{x})) \geq 1 - \alpha$.*

The computational savings of this sequential strategy are large. For example, assume that we want to find the positive level u excursion set $E_{u,\alpha}^+(x)$, and have m candidates $D(\rho_1),\ldots,D(\rho_m)$ to choose from. Using the naïve optimization method, we would then have to check whether $\mathsf{P}(D(\rho_i) \subseteq A_u^+(x)) > 1 - \alpha$ for each of these sets, and among the sets that satisfy the condition select the largest. Thus doing the probability calculation m times. However, by reordering the nodes and adding them sequentially we only have to run the integration routine once. Also note that the excursion function $F_u^+(\mathbf{s})$ is retrieved by setting $\alpha = 1$ and in each step saving the probabilities $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$. Thus, the entire excursion function can be retrieved by running the algorithm once.

Before extending this method to more general situations, we go into more detail on how to do the steps in Algorithm 3.1 in practice. In Section 3.1, a few sequential integration methods are presented. In Section 3.2, some different parametric families for the excursion sets and level avoidance sets are introduced and Algorithm 3.1 is extended using two-parameter families. The problem of how to optimally reorder the nodes is also discussed in this section. Finally in Section 3.3, three different methods are proposed for calculating excursion sets under the full posterior distribution (2).

3.1 Gaussian probability calculations

For a Gaussian vector $\mathbf{x}$, the probabilities $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$, $\mathsf{P}(D \subseteq A_u^-(\mathbf{x}))$, and $\mathsf{P}(D^+ \subseteq A_u^+(\mathbf{x}), D^- \subseteq A_u^-(\mathbf{x}))$ can all be written on the form

$$I(\mathbf{a}, \mathbf{b}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \int_{\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}} \exp(-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{3}$$

where $\mathbf{a}$ and $\mathbf{b}$ are vectors depending on the mean value of $\mathbf{x}$, the domain D, and on u. There have been considerable research efforts devoted to approximating integrals of this form in recent years, and we will in this section briefly describe a few techniques that can be used.

The simplest way of approximating (3) is to use Monte-Carlo (MC) integration. However, estimating the probability with any reasonable accuracy using standard MC integration is often too computationally expensive. Fortunately there are a number of variance reduction techniques that can be used to increase the efficiency.

A key step in many numerical integration techniques is to transform the integral to make it more suitable for integration. Notably, Genz [1992] derived such a transformation for the Gaussian integral (3), though similar transformations have been suggested by other authors as well [see e.g. Geweke, 1991]. Besides transforming the integral to the unit hyper cube, the transformation also achieves a separation of the variables so that the full problem can be calculated sequentially. The integral can then efficiently be evaluated using a quasi MC (QMC) method where the uniform random numbers in the ordinary MC integrator are replaced by some deterministic sequence of points chosen to reduce the probabilistic error bound of the crude MC integrator, see Genz and Bretz [2009] for details.

A final variance reduction technique for the general integration problem is to reorder the variables before calculating the integral, as first suggested by Schervish [1984] and later improved by Gibson et al. [1994]. These reorderings can reduce the error by an order of magnitude, as shown by Genz and Bretz [2002]. However, the technique will not be applicable in our situation since the reordering will be determined by a parametric family for the excursion sets.

3.1.1 Methods for Markov random fields

A common assumption in spatial statics and image analysis is that the latent field can be modeled, or approximated, using a Gaussian Markov random field (GMRF). See Rue and Held [2005] for an introduction to GMRFs, and note that GMRFs are used also for modeling in continuous space, for example using the SPDE approach by Lindgren et al. [2011]. One of the motivating reasons for using GMRFs is that it reduces the computational cost for parameter estimation and spatial prediction, and because of this one would also like to be able to use the Markov property in the calculation of (3).

The main difference between latent GMRF models and standard Gaussian models is that the distribution is specified using the (sparse) precision matrix $\mathbf{Q}$ instead of the covariance matrix:

$$I(\mathbf{a}, \mathbf{b}, \mathbf{Q}) = \frac{|\mathbf{Q}|^{1/2}}{(2\pi)^{d/2}} \int_{\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}} \exp(-\frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x}) \, d\mathbf{x}, \tag{4}$$

Using the QMC methods on GMRF models is difficult without first inverting the precision matrix and then ignoring the sparsity of $\mathbf{Q}$ in the calculations. To take advantage of the sparsity of $\mathbf{Q}$ one can use the fact that any GMRF can be viewed as a non-homogeneous auto-regressive process defined backwards in the indices of $\mathbf{x}$ [see Rue and Held, 2005, Theorem 2.7], that is, if $\mathbf{x}$ is a GMRF with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$, then

$$x_i | x_{i+1}, \ldots, x_n \sim \mathsf{N}\left(\mu_i - \frac{1}{L_{ii}} \sum_{j=i+1}^n L_{ji}(x_j - \mu_j), L_{ii}^{-2}\right), \tag{5}$$

where L_{ij} are the elements of the Cholesky factor of $\mathbf{Q}$. The integral can thus be written as

$$I(\mathbf{a}, \mathbf{b}, \mathbf{Q}) = \int_{a_1}^{b_1} \pi(x_1 | \mathbf{x}_{2:d}) \int_{a_2}^{b_2} \pi(x_2 | \mathbf{x}_{3:d}) \cdots \int_{a_{d-1}}^{b_{d-1}} \pi(x_{d-1} | x_d) \int_{a_d}^{b_d} \pi(x_d) \, d\mathbf{x}$$

where, because of the Markov structure, $x_i | \mathbf{x}_{i+1:d}$ only depends on the elements in $x_{\mathcal{N}_i \cap \{i+1:d\}}$, and $\mathcal{N}_i$ is the neighborhood of i in the graph of the GMRF.

If $\mathbf{Q}$ is a band-matrix, the integral can be efficiently calculated as a sequence of iterated one-dimensional integrals as discussed in Genz and Kahaner [1986]. However, the band width of $\mathbf{L}$ will often be too large for this method to be efficient, and a better alternative is then of use a particle filter algorithm based on the GHK simulator [Geweke, 1991, Hajivassiliou, 1991, Keane, 1993]. Denote the integral of the last $d - i$ components by I_i,

$$I_i = \int_{a_i}^{b_i} \pi(x_i | \mathbf{x}_{i+1:d}) \cdots \int_{a_{d-1}}^{b_{d-1}} \pi(x_{d-1} | x_d) \int_{a_d}^{b_d} \pi(x_d) \, d\mathbf{x},$$

and note that the integral is the normalizing constant to the truncated density

$$f_i(\mathbf{x}_{i:d}) \propto 1(\mathbf{a}_{i:d} < \mathbf{x}_{i:d} < \mathbf{b}_{i:d})\pi(\mathbf{x}_{i:d}).$$

The integrals $I_d, \ldots, I_1$ are now estimated sequentially using importance sampling. In the first step, calculate $I_d = \Phi(L_{ii}(b_d - \mu_d)) - \Phi(L_{ii}(a_d - \mu_d))$, simulate N samples $\{x_d^j\}_{j=1}^N$ from the truncated normal distribution $h_d(x_d) \propto 1(a_d < x_d < b_d)\pi(x_d)$, and set $w_d^j = I_d$. Next, simulate x_{d-1}^j from

$h_{d-1}(x_{d-1}|x_d^j) = 1(a_{d-1} < x_{d-1} < b_{d-1})\pi(x_{d-1}|x_d^j)$ and set $\mathbf{x}_{d-1:d}^j = \{x_{d-1}^j, x_d^j\}$. The integral I_{d-1} is estimated as $I_{d-1} \approx \frac{1}{N}\sum_{j=1}^N w_{d-1}^j$ where w_{d-1}^j are the importance weights

$$w_{d-1}^j = \frac{1(\mathbf{a}_{d-1:d} < \mathbf{x}_{d-1:d} < \mathbf{b}_{d-1:d})\pi(\mathbf{x}_{d-1:d})}{h_{d-1}(\mathbf{x}_{d-1}|\mathbf{x}_d)h_d(\mathbf{x}_d)}. \tag{6}$$

Proceed like this, simulating from the truncated conditional distributions and in each step updating the importance weights recursively through

$$w_i^j = \left[\Phi\left(L_{ii}(b_i - \mu_i) + \sum_{j=i+1}^n L_{ji}(X_j - \mu_j) \right) - \Phi\left(L_{ii}(a_i - \mu_i) + \sum_{j=i+1}^n L_{ji}(X_j - \mu_j) \right) \right] w_{i+1}^j.$$

To reduce the variance of the estimator when the target probability is small, a resampling step can be performed after having calculated the weights w_i^j. This is a common strategy in particle filter techniques, and the sample $\{\mathbf{x}_{i:d}^j\}$ is then updated by selecting N particles from the set, where $x_{i:d}^j$ is selected with probability $w_i^j / \sum_{k=1}^N w_i^k$. To avoid resampling too often, one can do the resampling only if some criterion is met, for example if the effective sample size is below some given threshold.

3.2 Parametric families

In theory one can use any shape optimization technique to find the largest region D. However, since evaluating the probability $\mathsf{P}(x(\mathbf{s}) > u, \mathbf{s} \in D)$ for a given set D is computationally expensive, one would like to do as few iterations as possible in this step. As discussed previously, we will solve this by assuming a parametric form of the sets D. The optimization can then be reduced to a standard optimization of only a few variables instead of doing a full shape optimization procedure. The parametric families will be based on the marginal quantiles of $x(\mathbf{s})$,

$$\mathsf{P}(x(\mathbf{s}) \leq q_\rho(\mathbf{s})) = \rho,$$

which are easy to calculate using only the marginal posterior distributions. The simplest one-parameter family based on the marginal quantiles is given in the following definition.

Definition 3.2 (One-parameter family). Let $q_\rho(\mathbf{s})$ be the marginal quantiles for $x(\mathbf{s})$, then a one-parameter family for the positive and negative u excursion sets is given by

$$D_1^+(\rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) > u) \geq 1 - \rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \leq u) \leq \rho\} = A_u^+(q_\rho),$$
$$D_1^-(\rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) < u) \geq 1 - \rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \geq u) \leq \rho\} = A_u^-(q_{1-\rho}).$$

Using this one-parameter family in Algorithm 3.1 is equivalent to finding a threshold value for the marginal excursion probabilities to get the correct simultaneous significance level. It is thus similar to the thresholding algorithms discussed in Marchini and Presanis [2003] but with the important difference that the correct joint, often non-stationary, posterior density is used when finding the threshold.

The simple one-parameter family can be extended in a number of ways, for example by considering other levels in the excursion sets.

Definition 3.3 (Two-parameter family). Let $q_\rho(\mathbf{s})$ be the marginal quantiles for $x(\mathbf{s})$, then a two-parameter family for the positive and negative u excursion sets is given by

$$D_1^+(v, \rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) > v) \geq 1 - \rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \leq v) \leq \rho\} = A_v^+(q_\rho),$$
$$D_1^-(v, \rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) < v) \geq 1 - \rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \geq v) \leq \rho\} = A_v^-(q_{1-\rho}).$$

The sets $D_1^+(v, \rho)$ and $D_1^-(v, \rho)$ are increasing in ρ for a fixed v.

One drawback with this parametric family is that it does not take the spatial dependency of the data into account directly. Therefore certain sets which might seem reasonable to test are not included in the family. Consider the example in Section 4.1, Figure 1, Panel (a), where the marginal excursion probabilities are shown in grey for an example in one dimension where the model is a Gaussian process with exponential covariance function. The estimated posterior mean in this example is shown as the black curve in Panel (b) in the figure, and in this situation a reasonable candidate for the 0-excursion set might be a contiguous set centered at 1, $[1 - \lambda_1, 1 + \lambda_2]$ for some positive λ_1, λ_2. However, looking at the marginal probabilities we see that sets on this form will not be included in the parametric family. One way of including such sets is to first smooth the marginal excursion probabilities $p_i = \mathsf{P}(x(\mathbf{s}_i) > u)$ using some parametric smoother and then consider sets on the form $\{\mathbf{s}; p_i^\tau \geq 1 - \rho\}$ where p_i^τ are the smoothed positive excursion probabilities.

Definition 3.4 (Two-parameter smoothing family). Let p_i^τ be the smoothed marginal positive u excursion probabilities, using a circular averaging filter with radius τ. A two-parameter family for the positive and negative u excursion sets is then given by

$$D_2^+(\tau, \rho) = \{\mathbf{s}; p_i^\tau \geq 1 - \rho\},$$
$$D_2^-(\tau, \rho) = \{\mathbf{s}; p_i^\tau \leq \rho\}.$$

The parameter τ determines how close p_i^τ is to the original excursion probabilities. For $\tau = 0$, no smoothing is done and for a general τ, p_i^τ is equal to the average of the marginal excursion probabilities in the disk with radius τ centered at $\mathbf{s}_i$. As τ increases p_i^τ becomes smoother and approaches a constant function equal to the average excursion probability. One could also use other types of parametric smoothers instead of the simple averaging filter.

Using the two-parameter families requires a modification to Algorithm 3.5, resulting in a slightly more computationally demanding method.

Algorithm 3.5 (Calculating excursion sets using a two-parameter family). *Assume that the model parameters $\boldsymbol{\theta}$ are known and that the posterior distribution $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is Gaussian. Further assume that $D(\nu, \rho)$ is a parametric family for the possible excursion sets, such that $D(\nu, \rho_1) \subseteq D(\nu, \rho_2)$ if $\rho_1 < \rho_2$ for a fixed ν. The following strategy is then used to calculate $E_{u,\alpha}^+$.*

1. *Choose a suitable (sequential) integration method for the problem.*

2. *Select a suitable one-dimensional optimization strategy.*

3. *Do optimization of the size of $D(\nu, \bullet)$ over ν:*

   - *For the current value of ν, reorder the nodes to the order they will be added to the excursion set when the parameter ρ is increased.*
   - *sequentially add nodes to the set D according to the ordering given above and in each step update the probability $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$. Stop as soon as this probability falls below $1 - \alpha$.*
   - *return the last set D for which $\mathsf{P}(D \subseteq A_u^+(\mathbf{x})) \geq 1 - \alpha$.*

4. *$E_{u,\alpha}^+$ is given by the largest set D found in the optimization over ν.*

The optimization can in this case be done using a Golden section search or a similar fast optimization procedure for one-dimensional problems. Algorithm 3.5 can also be used to estimate uncertainty regions for contour curves by using the following two-parameter family for the pair of level avoiding sets.

Definition 3.6 (Parametric family for level avoiding sets). Let $D_1^+(\rho_1)$ and $D_1^-(\rho_2)$ be given by Definition 3.2. A two-parameter family for the pair of level avoiding sets is obtained as $(D_1^+(\rho_1), D_1^-(\rho_2))$. A one-parameter family is obtained by requiring that $\rho_1 = \rho_2 = \rho$.

The one-parameter family in Definition 3.6 can be used in Algorithm 3.1 to estimate level avoiding sets and uncertainty regions for contour curves without having to use the more computationally expensive Algorithm 3.5.

3.2.1 Domain bounds and reorderings

In the case of a GMRF posterior, it is desirable to make the Cholesky factor of the precision matrix as sparse as possible, because it reduces the number of floating point calculations that have to be done and reduces the error of the estimator. Reordering the nodes according to a parametric family does not guarantee good sparsity of the Cholesky factor, but the reordering can be improved by finding upper and lower bounds for the region.

The simplest upper bound for the region is to use

$$U_1 = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) \geq 1 - \alpha\},$$

which is calculated using only the marginal probabilities, and which is the largest region D if $x(\mathbf{s})$ is a perfectly correlated field. The domain D cannot contain any locations $\mathbf{s}$ which are not in U_1 because all points not in U_1 have marginal probabilities lower than $1 - \alpha$ of exceeding the level u.

A simple lower bound for the region is obtained using Boole's inequality as

$$L_1 = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) \geq 1 - \alpha/n\}$$

where n is the number of points in the discretization of the domain. In terms of multiple hypothesis testing, this lower bound is obtained from the classical Bonferroni correction method and an improved lower bound can be obtained using the Holm-Bonferroni method [Holm, 1979] as

$$L_2 = \{\mathbf{s} : p_{(k)} > 1 - \alpha/k\}$$

where $p_{(k)}$ is the kth largest probability in the set $\{\mathsf{P}(x(\mathbf{s}_i) > u), i = 1, \ldots, n\}$. If the stochastic variables $x(\mathbf{s}_i)$ are independent, L_2 is the largest domain D. If the variables are not independent or perfectly correlated, one has $L_2 \subset D \subset U_1$.

The nodes can now be categorized into three classes, the first class contains the nodes included in the lower bound L_2, the second class contains the nodes in the set $U_1 \setminus L_2$ and the third class contains all other nodes. Since one knows that all nodes in L_2 will be included in D, these can be reordered to maximize the sparsity of the Cholesky factor, for example using an approximate minimum degree permutation. The nodes in the second class are then added in the order determined by the parametric family. Finally, since the nodes in the third class will not be included in the domain, these can be reordered to maximize the sparsity or integrated out of the posterior distribution. Making the bounds more precise will improve the sparsity of the problem and therefore reduce the Monte-Carlo error and the computational complexity.

3.3 Probability calculations for the latent Gaussian setting

In practice, we cannot use the computations from the previous sections directly unless we are in a purely Gaussian setting with known parameters. In the latent Gaussian setting with posterior (2), the method has to be modified. Since this is a latent Gaussian setting, Integrated Nested Laplace Approximations (INLA) [Rue et al., 2009] is a good alternative for estimating the posterior distributions $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$; however, the choice of method to use for estimating these distributions does not affect the excursion set estimation so any appropriate method can be used to estimate these distributions. Given estimates of the posterior distributions, we propose three methods for calculating the excursion probabilities, assuming that $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is Gaussian:

EB: (Empirical Bayes) Ignore the parameter uncertainty and calculate the probability conditionally on a parameter estimate. That is, estimate the excursion sets under the conditional posterior $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_0)$ where $\boldsymbol{\theta}_0$ for example is the maximum a posteriori estimate or the maximum likelihood estimate of $\boldsymbol{\theta}$.

QC: (Quantile Correction) Do a correction to the Gaussian probability calculations based only on the marginal posteriors in the following way. For each i use the marginal posterior to calculate $\mathsf{P}(x_i > a_i|\mathbf{y})$ and $\mathsf{P}(x_i < b_i|\mathbf{y})$ and calculate $\tilde{a}_i$ and $\tilde{b}_i$ so that $\mathsf{P}(z_i > \tilde{a}_i|\mathbf{y}, \boldsymbol{\theta}_0) = \mathsf{P}(x_i > a_i|\mathbf{y})$ and $\mathsf{P}(z_i < \tilde{b}_i|\mathbf{y}, \boldsymbol{\theta}_0) = \mathsf{P}(x_i < b_i|\mathbf{y})$. An estimate of the probability is then given by $I(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \mathbf{Q}(\boldsymbol{\theta}_0))$, where $\mathbf{Q}(\boldsymbol{\theta}_0)$ is the posterior precision matrix for $\mathbf{x}$ given the estimated parameters.

NI: (Numerical Integration) Numerically approximate the excursion probability by approximating the integral in (2) as

$$\mathsf{P}(\mathbf{a} < \mathbf{x} < \mathbf{b}|\mathbf{y}) = \mathsf{E}(\mathsf{P}(\mathbf{a} < \mathbf{x} < \mathbf{b}|\mathbf{y}, \boldsymbol{\theta})) \approx \sum_{i=1}^{k} w_i \mathsf{P}(\mathbf{a} < \mathbf{x} < \mathbf{b}|\mathbf{y}, \boldsymbol{\theta}_i)$$

where the configuration of the points $\boldsymbol{\theta}_i$ in the hyper parameter space can, for example, be chosen as in the INLA method [see Rue et al., 2009] and the weights w_i are chosen proportional to $\pi(\boldsymbol{\theta}_i|\mathbf{y})$.

The EB method is the simplest, and may be sufficient in many situations. The QC method is based on correcting the limits of the integral so that the probability would be correct if the x_i's were independent. This method is as easy to implement as the EB method and should perform better in most scenarios. Finally, the NI method is k times more computationally demanding as the probability has to be calculated for each parameter configuration $\boldsymbol{\theta}_i$, but should also be the most exact method. If the number of parameters is small one can often obtain accurate results with only a few parameter configurations, but the accuracy of the estimator will depend on how these configurations are chosen.

A second modification is required if the conditional posterior $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_0)$ is not Gaussian. The simplest solution to this problem is to do a Gaussian approximation $\tilde{\pi}_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_0)$, for example using Laplace approximations or simplified Laplace approximations as suggested by Rue et al. [2009]. If a Gaussian approximation is not sufficient, the sequential integration method has to be modified, and how to do this will depend on the posterior distribution. For example, Genz and Bretz [2009] outline how the quasi Monte Carlo methods can be extended to t-distributions, and the GHK-based particle filter method can be extended to other types of distributions as well. However, it is outside the scope of this article to go into details on how to handle the non-Gaussian cases and we therefore leave this for future research.

4 Tests on simulated data

In this section, three examples using simulated data are presented to illustrate the methods and test their accuracy. In the first example, we look at a problem in one dimension with known model parameters, where a latent Gaussian process with an exponential covariance is observed under Gaussian measurement noise. In the second example, we compare the different parametric families for contour uncertainty sets for a model in two dimensions with known parameters, where a latent Gaussian Matérn field is observed under Gaussian measurement noise. In the third example, the same spatial model setup is used, but this time the model parameters are estimated from data and the three methods for handling the full posterior distribution are compared.

4.1 Example 1: 1d Gaussian data with known parameters

We begin with a simple one-dimensional example to illustrate the different sets we have previously defined. Let $x(s)$, $s \in [0, 2]$ be a Gaussian process with an exponential covariance function with scaling parameter $\lambda = 1$ and mean

$$\mu(s) = \begin{cases} s - 0.5 & \text{if } s < 1 \\ 1.5 - s & \text{if } s \geq 1. \end{cases}$$
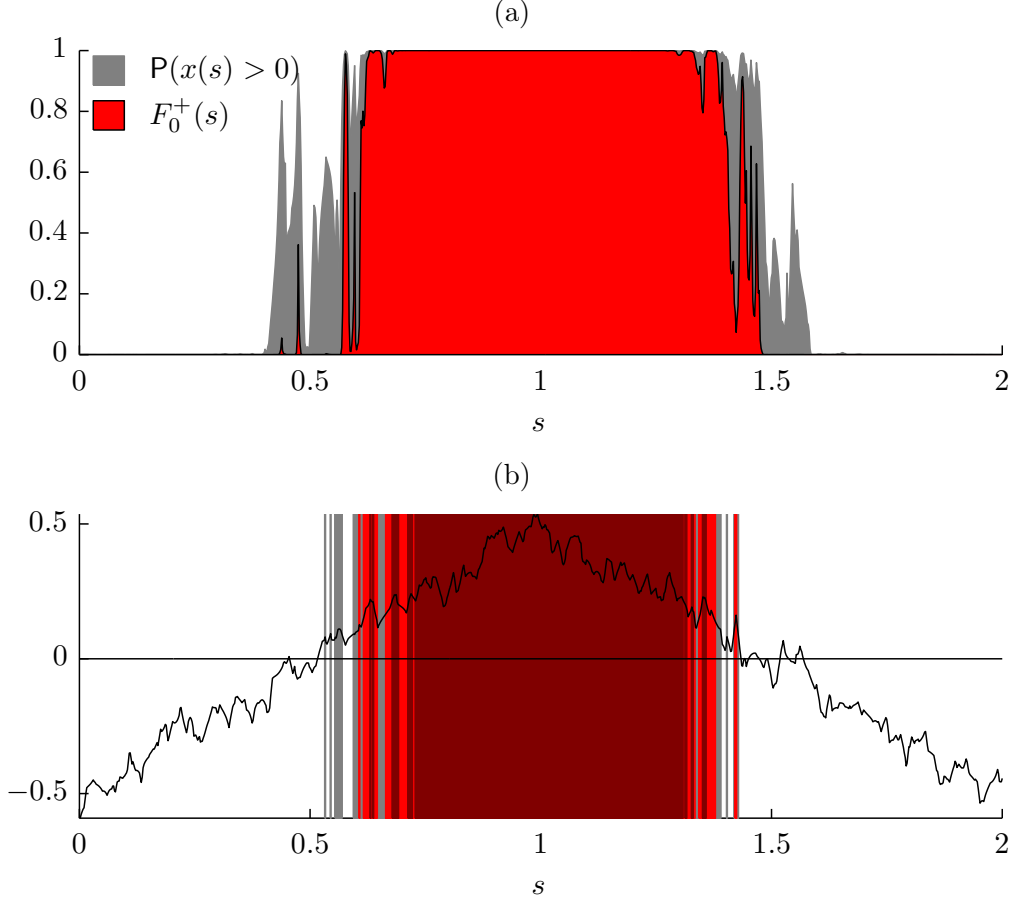
Figure 1: Results from Example 1. Panel (a) shows the excursion function $F_0^+(s)$ (red) and the marginal excursion probabilities $p(s) = \mathsf{P}(x(s) > 0)$ (grey). Panel (b) shows $E_{0,0.05}^+(x)$ in red, obtained as $A_{0.95}^+(F_0^+)$. The grey area shows $A_{0.95}^+(p)$, which is the upper bound U_1, and the dark red set is the lower bound L_2. The black curve is the kriging estimate of $x(s)$.

We generate a trajectory from the model and observe it at 500 locations $s_1, \ldots, s_n$ drawn at random in the interval under Gaussian measurement noise, giving us observations $y_i = x(s_i) + \varepsilon_i$ where $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$. We do spatial prediction (kriging) to 1000 equally spaced locations in the interval given the parameters $\boldsymbol{\theta}$ and the measurements $\mathbf{y}$, and then estimate the positive 0-excursion function $F_0^+(s)$ using the parametric family $D_1^+(0, \rho)$. In Figure 1, Panel (a), $F_0^+(s)$ is shown in red together with the marginal excursion probabilities $\mathsf{P}(x(s) > 0)$ in grey.

By the definition of $F_0^+(s)$, the positive 0-excursion set $E_{0,\alpha}^+(x)$, is obtained by calculating the $1 - \alpha$ excursion set of the function $F_0^+(s)$, and this set is shown for $\alpha = 0.05$ in red in Figure 1, Panel (b). The grey set shows the upper bound U_1, which is the set where $\mathsf{P}(x(s) > 0) \geq 1 - \alpha$, and the dark red set shows the Holm-Bonferroni lower bound L_2. The black curve shows the kriging estimate of the process given the data. Note that the grey and red sets are obtained as excursion sets of the grey and red functions in Panel (a), and also note that $L_2 \subset E_{0,\alpha}^+(x) \subset U_1$.

We now want to verify that the estimated sets $E_{0,\alpha}^+(x)$ have the correct excursion probability, that is, that $\mathsf{P}(x(s) > 0, s \in E_{0,\alpha}^+(x)) = 1 - \alpha$. To that end, draw N samples, $x_1(s), \ldots, x_N(s)$ from $\pi(x|\mathbf{y}, \boldsymbol{\theta})$, count the number of samples for which $\inf\{x(s), s \in E_{0,\alpha}^+(x)\} \geq 0$, and denote this number by N_s. Further let $\hat{p}(\alpha)$ denote the proportion of samples, N_s/N, that satisfies the requirement. If $E_{0,\alpha}^+(x)$ is correctly estimated, $\hat{p}(\alpha)$ should be close to $1 - \alpha$. In Figure 2, Panel (a), the difference $1 - \alpha - \hat{p}(\alpha)$ is shown as a function of $1 - \alpha$. The difference is calculated twice, using two different estimates $\hat{p}(\alpha)$, each based on $N = 50000$ samples. As can be seen in the figure, the difference is very
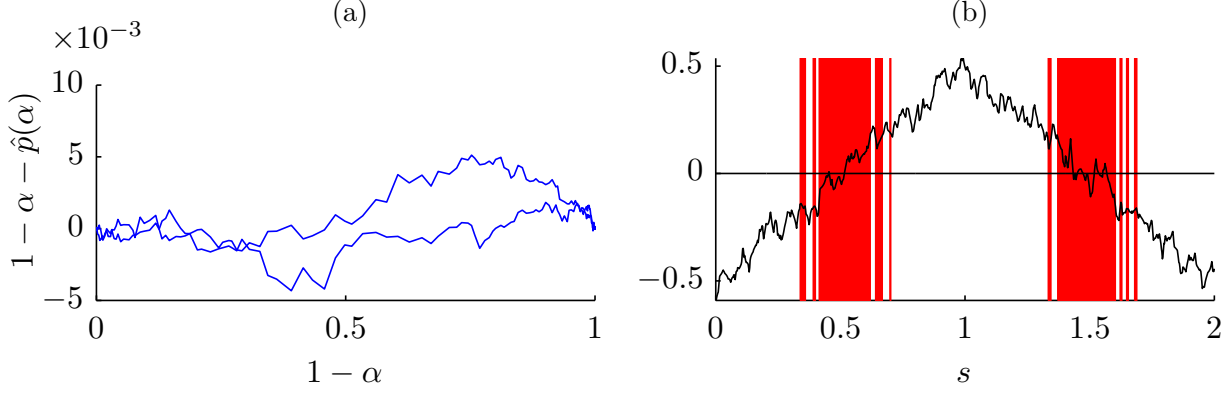
Figure 2: Results from Example 1. Panel (a) shows two estimates of $1 - \alpha - \hat{p}(\alpha)$ as a function of $1 - \alpha$. $\hat{p}(\alpha)$ is an estimate of $\mathsf{P}(x(\mathbf{s}) > 0, \mathbf{s} \in E_{0,\alpha}^+(x))$ based on Monte-Carlo simulation of $x(\mathbf{s})$, which should be close to $1 - \alpha$ if $E_{0,\alpha}^+(x)$ is correctly estimated. The two curves show two results for two different Monte-Carlo simulations. Panel (b) shows the estimated contour uncertainty set $M_{0,0.05}^c(x)$.

small for all values of α, and the difference that can be seen is mostly due to the Monte-Carlo error in the estimation of $\hat{p}(\alpha)$, which has nothing to do with the accuracy of the method. Thus the sets $E_{0,\alpha}^+$ indeed have the correct excursion probabilities.

Finally in Figure 2, Panel (b), the 0-contour uncertainty region $M_{0,0.05}^c(x)$ is shown in red and the kriging estimate of $x(s)$ is again shown in black. The set was estimated using the two-parameter family for level avoidance sets from Definition 3.6 and Algorithm 3.5. The complement of this set is the union of the level avoiding sets $(M_{0,0.05}^-(x), M_{0,0.05}^+(x))$, which is the largest pair of sets (D^+, D^-) satisfying $\mathsf{P}(D^- \subseteq A_u^-(x), D^+ \subseteq A_u^+(x)) \geq 0.95$.

4.2 Example 2: 2d Gaussian data with known parameters

In this example, we change to a spatial model to test the parametric families for contour sets. Let $x(\mathbf{s})$, $\mathbf{s} \in [0, 10] \times [0, 10]$, be a Gaussian field with a constant mean $\mu = 0$ and a Matérn covariance function

$$C(\|\mathbf{h}\|) = \frac{2^{1-\nu}\phi^2}{(4\pi)^{\frac{d}{2}}\Gamma(\nu + \frac{d}{2})\kappa^{2\nu}}(\kappa\|\mathbf{h}\|)^\nu K_\nu(\kappa\|\mathbf{h}\|), \tag{7}$$

where ν is a shape parameter, κ^2 a scale parameter, ϕ^2 a variance parameter, K_ν is a modified Bessel function of the second kind of order $\nu > 0$, and $\|\cdot\|$ denotes the Euclidean spatial distance. We use the SPDE representation by Lindgren et al. [2011] of the field using a triangulation based on an 80×80 regular lattice in the region. The representation is a piecewise linear approximation $x(\mathbf{s}) \approx \sum_i x_i \varphi_i(\mathbf{s})$ of the field using 6400 piecewise linear functions $\varphi_i(\mathbf{s})$, each centered at one of the nodes in the lattice. The advantage with this representation is that it allows us to do all calculations using the weights $\mathbf{x}$ of the basis expansion, which form a Gaussian Markov random field.

We set $\nu = \phi = 1$, and $\kappa^2 = 0.5$, and generate a sample of the field and measure it at 1000 locations in the square, chosen at random, under Gaussian measurement noise, giving us observations $y_i = x(\mathbf{s}_i) + \varepsilon_i$ where $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$ and $\sigma = 0.1$. The posterior estimate (kriging) of $\mathbf{x}|\mathbf{y}$ can be seen in Figure 3, Panel (a), and the uncertainty region $M_{0,0.05}^c(x)$ for the 0-contour can be seen in Panel (b). In this case the $M_{0,0.05}^c(x)$ was estimated using the one-parameter family in Definition 3.6, and it is now of interest to test how much is gained by using the two-parameter family from the same definition instead.

To that end, we generate 50 data sets using the same setup, and for each data set estimate $M_{0,0.05}^c(x)$, first using the one-parameter family $(D_1^+(\rho), D_1^-(\rho))$, and then using the more general two-parameter family $(D_1^+(\rho_1), D_1^-(\rho_2))$. Since the one-parameter family is a special case of the two-parameter family where $\rho_1 = \rho_2 = \rho$, the contour sets estimated with the two-parameter family
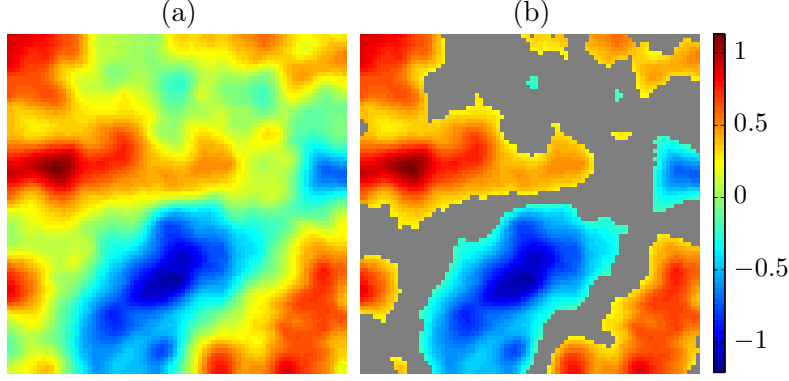
13

Figure 3: Results from Example 2. Panel (a) shows a kriging estimate and Panel (b) shows the same estimate where the corresponding estimated contour uncertanty set $M_{0,0.05}^c(x)$ is superimposed in grey.

should always be smaller than the one-parameter sets. However, using the two-parameter family, the estimated sets are on average only 0.2% smaller than if the one-parameter family is used, so in this case it is arguably not worth the extra computational effort to use the two-parameter family, although for other levels u, or other latent models, the difference might be larger.

4.3 Example 3: 2d Gaussian data with unknown parameters

In this example we compare the three methods, described in Section 3.3, for handling the full posterior distribution (2) in the calculations. The same Gaussian Matérn model is used as in Example 2, with the difference that we now also estimate the parameters from the data.

We set $\nu = \phi = 1$, and $\kappa^2 = 2$ in the covariance function (7), and generate a sample of the field and measure it at 1000 locations in the square, chosen at random, under Gaussian measurement noise, giving us observations $y_i = x(\mathbf{s}_i) + \varepsilon_i$ where $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$ and $\sigma = 0.5$. Given the measurements, we estimate the parameters and the marginal posterior distributions using INLA. The posterior estimate (kriging) of $\mathbf{x}|\mathbf{y}$ can be seen in the lower right panel of Figure 4, and in the lower left panel, the marginal probabilities $\mathsf{P}(x(\mathbf{s}) > 0|\mathbf{y})$ are shown.

We now estimate the excursion function $F_0^+(\mathbf{s})$ using the three different methods described in Section 3.3 and the one-parameter family from Definition 3.2 for the excursion sets. These can be seen in the upper panels of Figure 4. Visually it is in this case difficult to see any differences between the three estimates of the excursion function. To compare the accuracy of the estimates we will do a simular comparison to the one performed in Example 1, where Monte-Carlo simulation was used to estimate $\hat{p}(\alpha)$, the proportion of samples satisfying $\inf\{x(\mathbf{s}), \mathbf{s} \in E_{0,\alpha}^+(x)\} \geq 0$, which should be close to $1 - \alpha$ if $E_{0,\alpha}^+(x)$ is correct.

There are three possible sources of errors in this comparison. The first one is the Monte-Carlo error from the estimation of $\hat{p}(\alpha)$, which has nothing to do with the accuracy of the method. The second error is the Monte-Carlo error in the probability estimation when estimating the excursion distribution functions. This error is, however many orders of magnitude smaller in this case. The final error is the approximation error induced by using any of the three methods EB, QC, or NI for handling the full posterior distribution.

To investigate this approximation error, the difference $1 - \alpha - \hat{p}(\alpha)$ is estimated for the three estimates of $F_0^+(\mathbf{s})$. First we base the estimate on 20000 samples from the posterior $\pi(\mathbf{x}|\mathbf{y})$, obtained using the MCMC sampler described in Appendix A. In Figure 5, Panel (a), the results can be seen for the EB method (blue), the QC method (green), the NI method with $k = 45$ parameter settings (red), and the NI method with $k = 15$ parameter settings (cyan). The comparison was done twice, with two different samples of size 20000 when calculating $\hat{p}(\alpha)$, and the curves of the same color show these two and give an indication of the size of the Monte-Carlo error in the comparison. As seen in
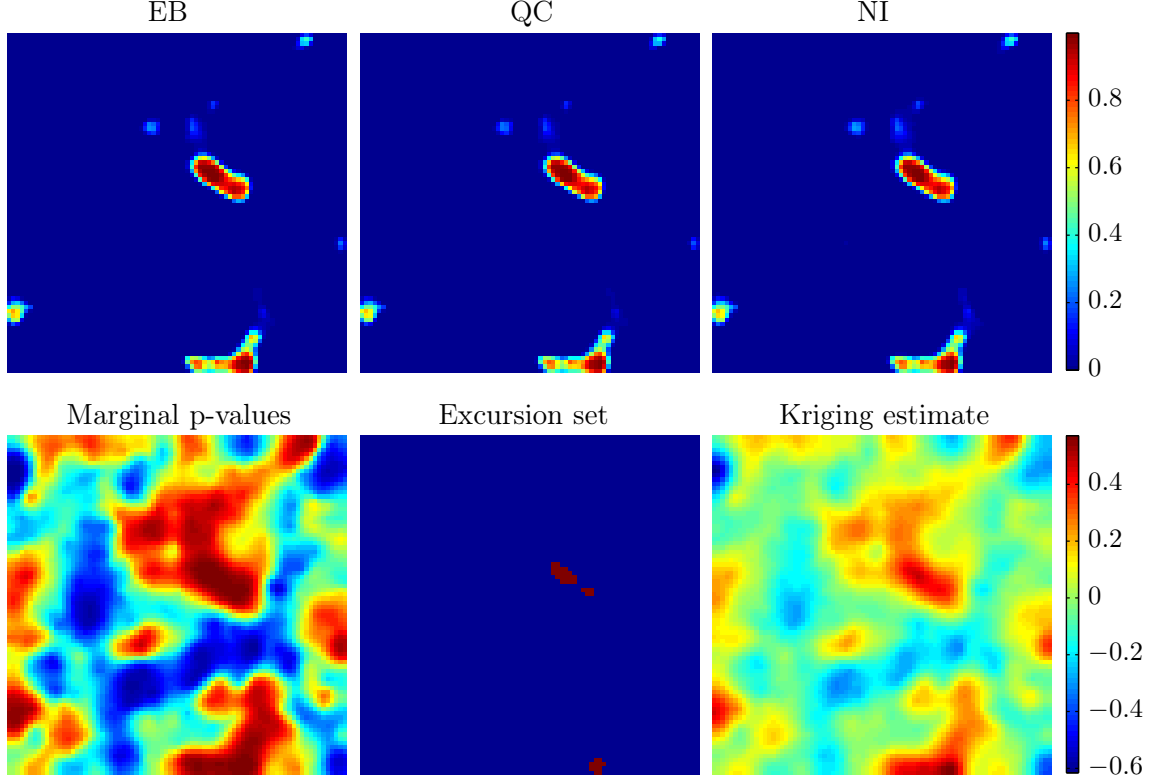
14

Figure 4: Results from Example 3. In the top row, three estimates of the excursion function can be seen using the EB method (left), the QC method (middle), and the NI method with 15 parameter configurations (right). In the bottom row, the marginal p-values for exceeding the limit can be seen in the left panel, using the same color scale as for the top row. The middle panel shows the set $E_{0,0.05}^+(x)$ given by excursion function estimated by the NI method. Finally the right panel shows the kriging estimate of the latent field.

the figure, the NI method performs best, as expected.

The error using the NI method comes from the fact that only finitely many points are used in the integration when approximating the posterior distribution for the parameters. That is, the full posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$ is approximated by a discrete distribution with point masses at the parameter configurations $\boldsymbol{\theta}_i$ used in the integration, $\pi(\boldsymbol{\theta}_i) = w_i$. To verify that this indeed is the source of the error in, we construct a second Monte-Carlo sampler where we instead of sampling $\boldsymbol{\theta}$ from the full posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$ sample it from the discrete distribution defined by the 45 parameter configurations used in the first NI method. Panel (b) in Figure 5 shows the same comparison as Panel (a) but where $\boldsymbol{\theta}$ is sampled from the discrete distribution. As expected, the error for the NI method with 45 parameter configurations is now smaller.

The Monte-Carlo error from estimating $\hat{p}(\alpha)$ is quite large in Figure 5, so to get a better understanding of the other errors, a larger study was also performed where the procedure in Figure 5 was repeated 50 times for 50 different simulated data sets, and for each data set $N = 60000$ draws from the posterior was used when estimating $\hat{p}(\alpha)$. The average errors of these 50 runs can be seen in Figure 6. In Panel (a) the results using samples from the full posterior is shown, and in Panel (b) the results using the discrete distribution for $\boldsymbol{\theta}$ is shown. Note that the red curve is very close to zero in Panel (b), indicating that the error in the NI method mostly depends on choosing the integration points for $\boldsymbol{\theta}$ so that they capture the true posterior distribution well. Also note that the QC method performs well for large values of $1 - \alpha$, and since one most often is interested in finding the excursion sets for small values of α, this method is then a good way of finding such sets with less computational effort than using the NI method.
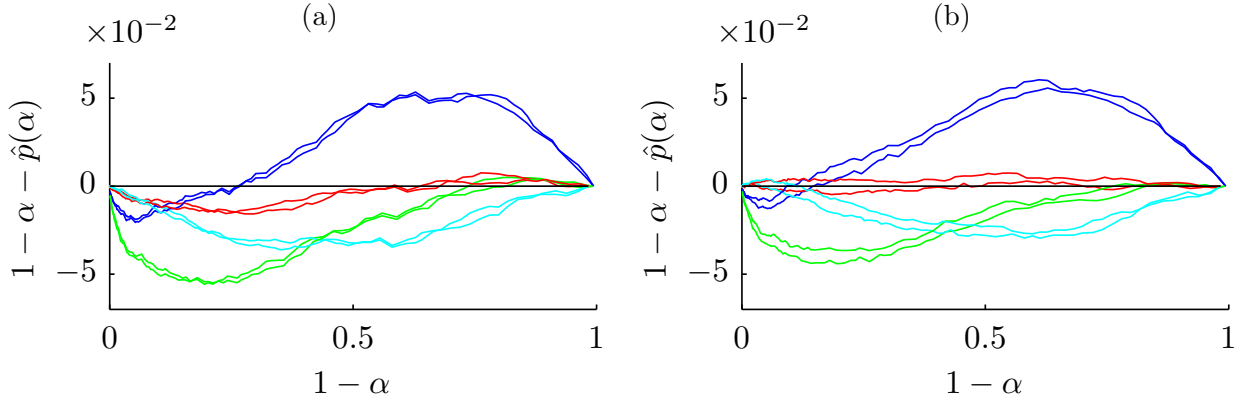
Figure 5: Results from Example 3 showing the difference $1 - \alpha - \hat{p}(\alpha)$ as a function of $1 - \alpha$ for the different approximation methods, EB (blue), QC (green), NI using 45 parameter configurations (red), and NI using 15 parameter configurations (cyan). $\hat{p}(\alpha)$ is an estimate of $\mathsf{P}(x(\mathbf{s}) > 0, \mathbf{s} \in E_{0,\alpha}^+(x))$ based on MCMC simulation of $x(\mathbf{s})$, which should be close to $1 - \alpha$ if $E_{0,\alpha}^+(x)$ is correctly estimated. In Panel (a), $\hat{p}(\alpha)$ is estimated using the full posterior distribution, and in Panel (b), $\hat{p}(\alpha)$ is estimated using the discrete posterior distribution defined by the 45 parameter configurations from the NI method. The comparison was done twice, with the two different estimates of $\hat{p}(\alpha)$, each based on 20000 samples of $x(\mathbf{s})$, and the curves of the same color shows these two.

5   Applications

In this section, we will use the techniques described above in two different applications. In the first, we study air pollution data from Piemonte region in northern Italy and estimate regions where the daily limit for PM_{10} (particulate matter with an aerodynamic diameter of less than 10 μm) is exceeded. In the second application, we study vegetation index data from the African Sahel and estimate areas that experienced a significant increase in vegetation after the drought period in the early 1980's. In both of these applications the data sets are large, and the Markov structure of the latent Gaussian models has to be used in the calculations.

5.1   Air pollution data

High levels of air pollution can be harmful for the ecosystems and the human health. The effects on human health ranges from minor effects to the cardio-respiratory system to premature mortality [Cohen et al., 2009, Cameletti et al., 2012]. Because of this, environmental agencies have to assess the air quality in order to take proper actions for improving the situation in polluted areas, and an important tool in this process is the ability to produce continuous maps of air pollution.

A region where the daily limit values fixed by the European Union for human health protection (see EU Council Directive 1999/30/EC) are periodically exceeded is the Piemonte region in northern Italy. Recently, Cameletti et al. [2012] proposed a statistical model to capture the complex spatio-temporal dynamics of PM_{10} concentration in the region and used it to produce daily maps of PM_{10}. They also produced daily maps of exceedance probabilities of the value $50\mu g/m^3$, which is the value fixed by the European directive 2008/50/EC for the daily mean concentration that cannot be exceeded more than 35 days in a year. These probability maps only considered the marginal excursion probabilities, and no attempts of producing maps of simultaneous exceedance probabilities were made. In the following, we will therefore consider the same model and data but also estimate the excursion functions for the $50\mu g/m^3$ limit value.

Cameletti et al. [2012] considers daily PM_{10} data measured at 24 monitoring stations by the Piemonte monitoring network during 182 days in the period October 2005 - March 2006, the data is provided by Aria Web Regione Piemonte. Denoting the measurements made at location $\mathbf{s}_i$ at time t
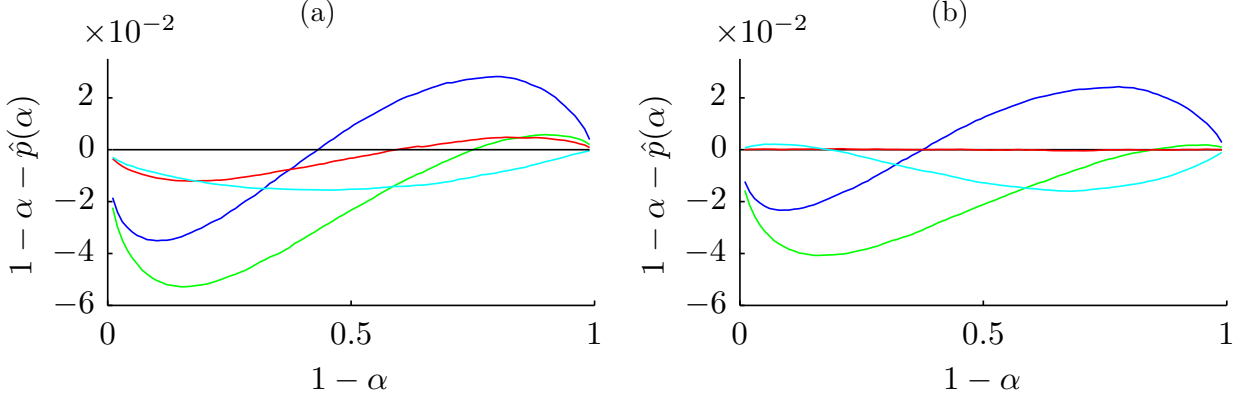
Figure 6: Results from Example 3 showing the difference $1 - \alpha - \hat{p}(\alpha)$ as a function of $1 - \alpha$ for the different approximation methods, EB (blue), QC (green), NI using 45 parameter configurations (red), and NI using 15 parameter configurations (cyan). The same setup as in Figure 5 was repeated 50 times for 50 different datasets and 60000 samples were used when estimating $\hat{p}(\alpha)$. This figure shows the average error of these 50 runs.

by $y(\mathbf{s}_i, t)$, the following measurement equation is assumed,

$$y(\mathbf{s}_i, t) = x(\mathbf{s}_i, t) + \varepsilon(\mathbf{s}_i, t), \tag{8}$$

where $\varepsilon(\mathbf{s}_i, t) \sim \mathsf{N}(0, \sigma_\varepsilon^2)$ is Gaussian measurement noise, both spatially and temporally uncorrelated, and $x(\mathbf{s}_i, t)$ is the latent field of true unobserved air pollutions. The latent field is assumed to be on the form

$$x(\mathbf{s}_i, t) = \sum_{k=1}^{p} z_k(\mathbf{s}_i, t)\beta_k + \xi(\mathbf{s}_i, t), \tag{9}$$

where the $p = 9$ covariates z_k are used and ξ is a spatio-temporal Gaussian random field. Based on the work of Cameletti et al. [2011] the following covariates were used: 1) Daily mean wind speed; 2) daily maximum mixing height; 3) daily precipitation; 4) daily mean temperature; 5) daily emissions; 6) altitude; 7) longitude; 8) latitude; and 9) intercept. These covariates are provided with hourly temporal resolution on a 4 km × 4 km regular grid by the environmental agency of Piemonte region (Arpa Piemonte), see Finardi et al. [2008]. The spatio-temporal process ξ is assumed to follow first order autoregressive dynamics in time with spatially dependent innovations:

$$\xi(\mathbf{s}_i, t) = a\xi(\mathbf{s}_i, t - 1) + \omega(\mathbf{s}_i, t), \tag{10}$$

where $|a| < 1$ and $\omega(\mathbf{s}_i, t)$ is a zero-mean temporally independent Gaussian process characterized by the spatio-temporal covariance function

$$\mathrm{Cov}(\omega(\mathbf{s}_i, t_1), \omega(\mathbf{s}_j, t_2)) = \begin{cases} 0 & \text{if } t_1 \neq t_2 \\ C(\|\mathbf{s}_i - \mathbf{s}_j\|) & \text{otherwise}, \end{cases} \tag{11}$$

where $C(\cdot)$ is a Matérn covariance function given by (7). The model parameters and the posterior distribution for the latent field are then estimated using INLA in combination with the SPDE representation of Lindgren et al. [2011], see Cameletti et al. [2012] for details.

The map of marginal excursion probabilities for the level $50\mu g/m^3$ for January 30, 2006, based on the estimated posterior distribution for x, can be seen in the left panel of Figure 7. To avoid inappropriate linear extrapolation of the effect of elevation beyond the range of the elevation of the observations, the results are only shown for areas below 1km. Based on these results, we now estimate the positive excursion function for the level $50\mu g/m^3$, $F_{50}^+(\mathbf{s})$, using the NI method from Section 3.3

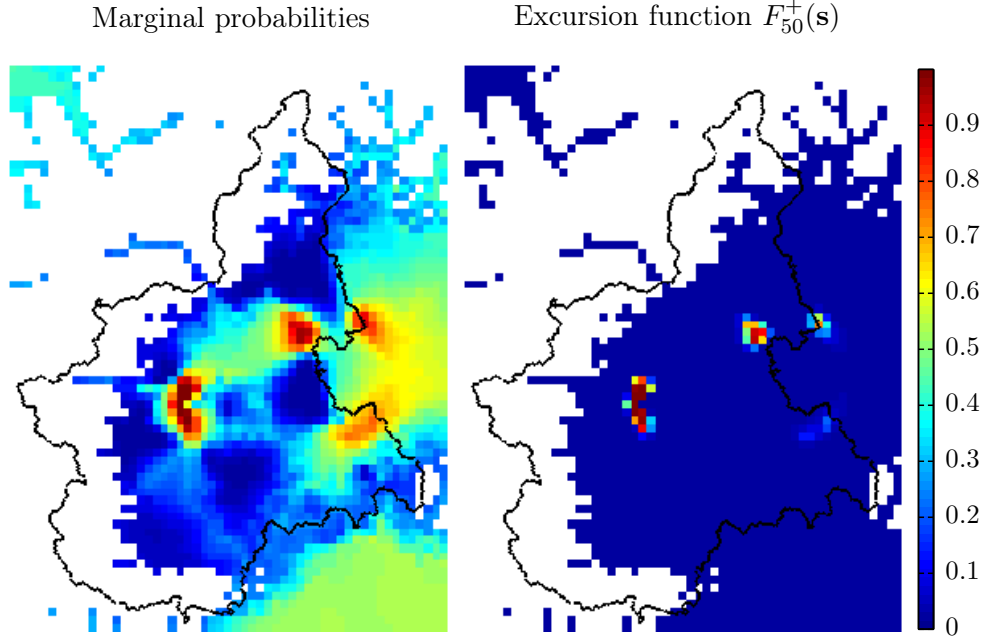Figure 7: Results from the PM_{10} application for January 30, 2006. A map of the marginal exceedance probabilities for $50\mu g/m^3$ (left), and the joint excursion distribution function for the level (right).

and the parametric family of excursion sets from Definition 3.2. A total of 25 parameter configurations are used in the integration. The result can be seen in the right panel of Figure 7. As seen in the figure, there are three regions where the level is clearly exceeded, and a fourth that possibly contains too high pollution levels. As expected, these areas coincide with the locations of the main metropolitan areas in the region; Turin, Novara, Vercelli, and Alessandria. In this case, it would have been desirable to make the predictions on a finer spatial scale, but since the covariates were given on a 4 km × 4 km grid, this spatial resolution had to be used in the prediction.

To get a better understanding of the results, it is also of interest to find the regions where the pollution level is simultaneously below the limit value with some given probability. The marginal probabilities for being below be the level $50\mu g/m^3$, based on the estimated posterior distribution for x, can be seen in the left panel of Figure 8. The results are again only shown for areas below 1km altitude. Using the same method as for the positive excursion function, we now estimate the negative excursion function for the level $50\mu g/m^3$, $F_{50}^-(\mathbf{s})$. The result can be seen in the right panel of Figure 8.

Note that the union of $E_{50,0.1}^+(x)$ and $E_{50,0.1}^-(x)$ covers only a small part of the region, indicating that the uncertainty in the problem is large. See the red and blue sets in the left panel of Figure 9. Also, by taking the complement of the set $E_{50,0.1}^-(x)$, we get the region that contains all exceedances of the level with certainty 0.9, indicated in grey in the left panel of Figure 9. This set is large, indicating that there are many regions where the level possibly is exceeded. Hence, it is important to note that the positive excursion set $E_{50,0.1}^+(x)$ is small because the uncertainty is large in the problem, and not because the other regions certainly have concentrations below the level.

To verify that the uncertainty is large, we finally calculate the contour function for the level $50\mu g/m^3$, $F_{50}^c(\mathbf{s})$, using the NI method and the one-parameter family from Definition 3.6 for the pair of level avoiding sets. The result can be seen in the right panel of Figure 9, and the uncertainty region for the contour curve indeed covers a large part of the region which indicates that the uncertainty in the estimated contour curve is large.
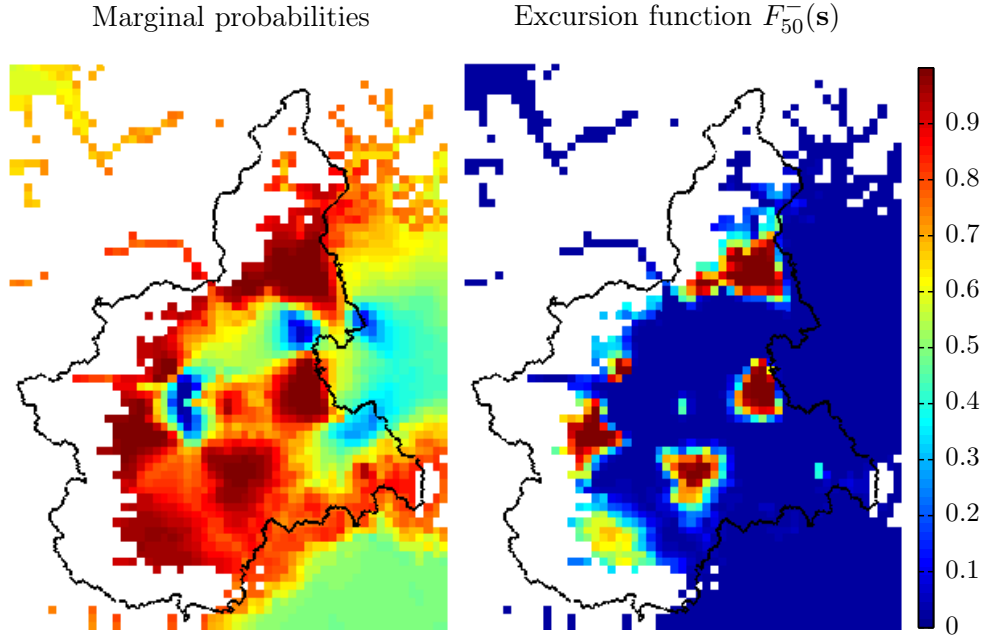
Figure 8: Results from the PM$_{10}$ application for January 30, 2006. A map of the marginal probabilities for the field being below the level $50\mu g/m^3$ (left), and the joint negative excursion distribution function for the level (right).

5.2 Spatially dependent temporal trends in vegetation data

Trends in vegetation cover are related to changes in climatic drivers, feedback mechanisms between the atmosphere and land surface, and human interaction. A region with rapid recent changes is the African Sahel. This zone has received much attention regarding desertification and climatic variations [Olsson, 1993, Nicholson, 2000, Lamb, 1982]. Recently, Eklundh and Olsson [2003] observed a strong increase in seasonal vegetation index over parts of the Sahel using Advanced Very High Resolution Radiometer (AVHRR) data from the NOAA/NASA Pathfinder AVHRR Land (PAL) database [Agbu and James, 1994, James and Kalluri, 1994], for the period 1982-1999. The study was based on ordinary least squares linear regression on individual time series extracted for each pixel in the satellite images. The results of Eklundh and Olsson [2003] were later improved by Bolin et al. [2009] where a spatial model for the vegetation was used in the analysis to capture the spatial dependencies in the trend estimation.

To find regions where changes in the vegetation have occurred over the course of the studied time period, both Eklundh and Olsson [2003] and Bolin et al. [2009] used significance testing for the individual pixels in the field. Thus, pixels that individually had significant changes in vegetation were found, but no attempts were made to find simultaneous excursion regions. Here, we will use a similar model to that of Bolin et al. [2009] but also estimate joint excursion regions for the vegetation trends.

Assume that the vegetation measurements year t are generated as,

$$\mathbf{Y}_t | \mathbf{X}_t, \mathbf{\Sigma}_{\epsilon_t} \in \mathsf{N}(\mathbf{A}_t \mathbf{X}_t, \mathbf{\Sigma}_{\epsilon_t}),$$

where $\mathbf{X}_t$ is the latent vegetation field with prior distribution $\pi(\mathbf{X}_t)$, $\mathbf{\Sigma}_{\epsilon_t}$ is a measurement noise covariance matrix, and $\mathbf{A}_t$ is an observation matrix determining which pixels in the field that are observed. To estimate time varying trends in the observations, $\mathbf{X}$ is restricted to follow a field of spatially varying linear trends:

$$\mathbf{X}_t = \mathbf{K}_1 + t\mathbf{K}_2 \tag{12}$$

The prior distribution for $\mathbf{K} = [\mathbf{K}_1^\top, \mathbf{K}_2^\top]^\top$ is obtained by evaluating the joint distribution for $\mathbf{X} = [\mathbf{X}_1^\top, \ldots, \mathbf{X}_T^\top]^\top$ conditionally on the restriction (12). We choose a second-order polynomial IGMRF
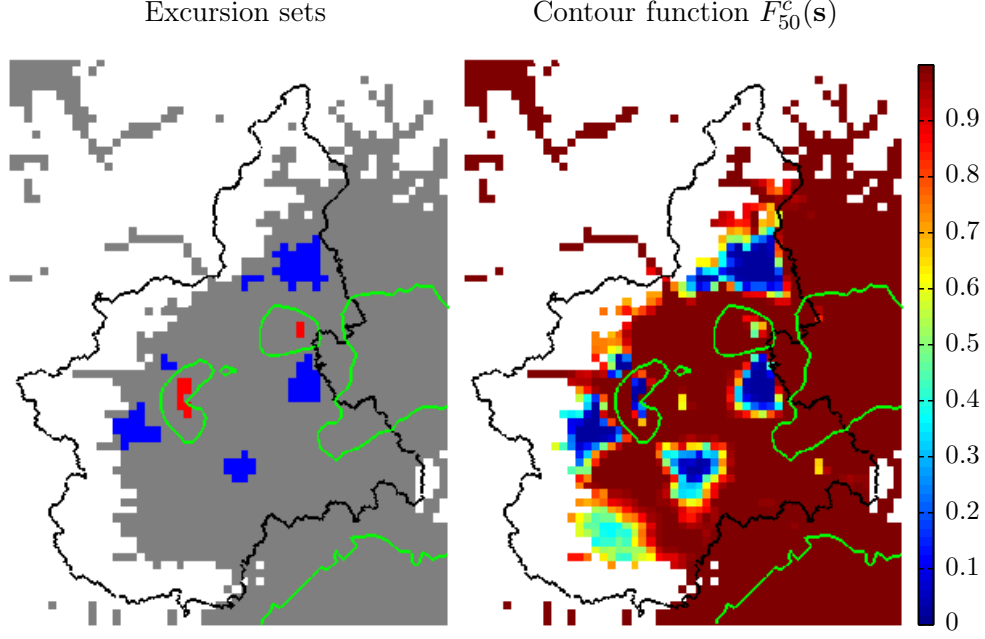
19

Figure 9: Results from the PM$_{10}$ application for January 30, 2006. In the left panel, the set $E^+_{50,0.1}(x)$ is shown in red, $E^-_{50,0.1}(x)$ in blue, and it's complement $E^+_{50,0.1}(x)^c$ in grey. The contour curve for the level $50\mu g/m^3$ is shown in green. The right panel shows the contour function for the level $50\mu g/m^3$, $F^c_{50}(s)$.

[Rue and Held, 2005, Section 3.4.2] prior for $\mathbf{X}$ and then calculate the corresponding prior distribution for $\mathbf{K}$, [see Bolin et al., 2009, for details].

To complete the model, the structure of $\mathbf{\Sigma}_\epsilon$ needs to be determined. Many of the factors that the measurement noise should model are local phenomena, such as aerosol and cloud cover. Since it seems unreasonable that the scale of these disturbances would be the same over the entire region, Bolin et al. [2009] assumed that the measurement noise was uncorrelated with a different noise variance at each pixel in the field. This results in a large number of parameters for the measurements noise, one for each pixel in the field, so here we instead use a different slightly simplified noise model. We divide the region into five different land cover categories using the Africa Land Cover Characteristics Data Base Version 2.0 (http://edc2.usgs.gov/glcc/glcc.php): *1*) Bare desert; *2*) Semi desert; *3*) Savanna; *4*) Crops, grass, and shrubs; and *5*) Forests and wetlands. The measurement noise variance at pixel $\mathbf{s}_i$ is then modeled as

$$\log \sigma^2(\mathbf{s}_i) = \sum_{k=1}^{5} \theta_k b_k(\mathbf{s}_i), \tag{13}$$

where $b_k(\mathbf{s})$ is the spatial basis function with values equal to the proportion of vegetation type k at each pixel $\mathbf{s}$. The parameters of the model are thus the scale parameter κ and the five measurement noise parameters $\theta_1, \ldots, \theta_5$. A gamma prior is assumed for κ and gaussian priors are used for θ_k.

We choose to study the western part of the Sahel region, and this area is divided into 35463 pixels of size 8 km $\times$ 8 km, so the field $\mathbf{K}$ has 70926 elements, and there are 547832 measurements from 17 years of data starting in 1982 and ending in 1999.

The model parameters and the marginal posterior distributions are estimated using the INLA framework and the excursion sets are estimated using the QC method from Section 3.3. The results can be seen in Figure 10 and Figure 11. The top panel in Figure 10 shows the posterior estimates of the intercepts, $\mathbf{K}_1$, and the slopes, $\mathbf{K}_2$, is shown in the bottom panel. As expected, intercepts are larger in the savanna regions to the south, and smaller in the semi desert regions to the north. The top panel of Figure 11 shows the estimated standard deviation of the measurement noise using model
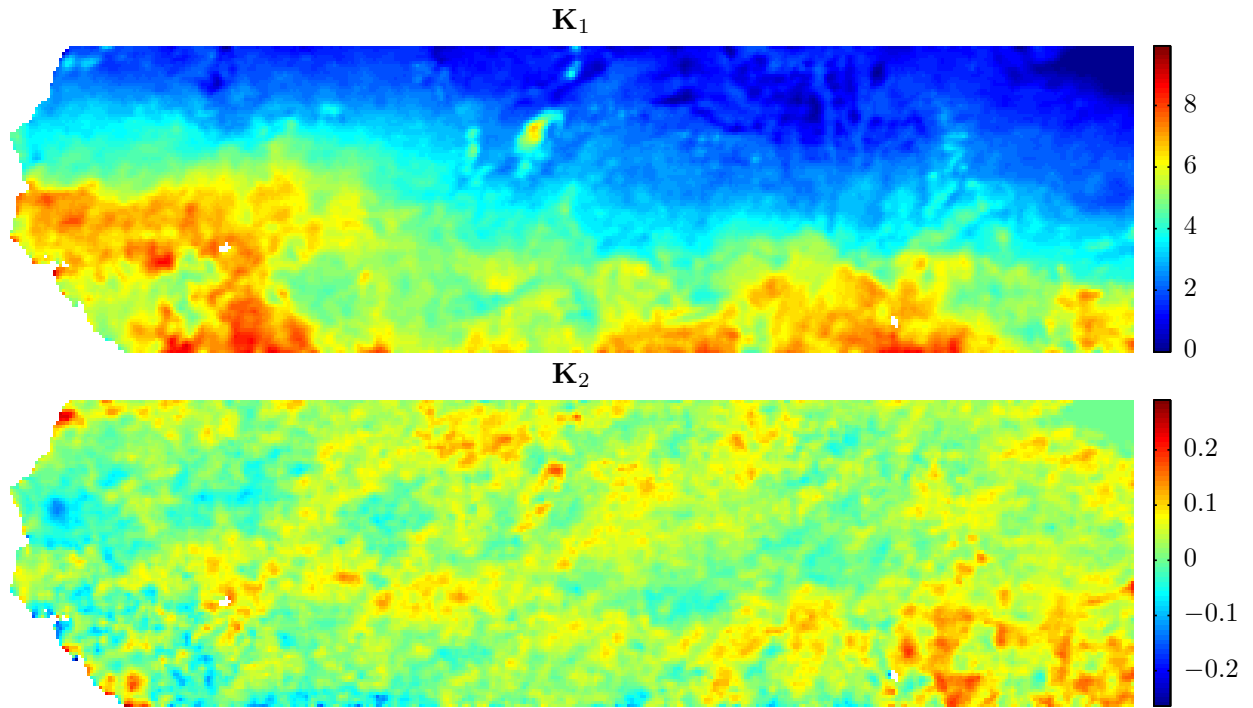
Figure 10: Results from the Sahel vegetation data. The top panel shows the posterior estimates of the regression intercepts, $\mathbf{K}_1$, and the bottom panel shows the estimated slopes, $\mathbf{K}_2$.
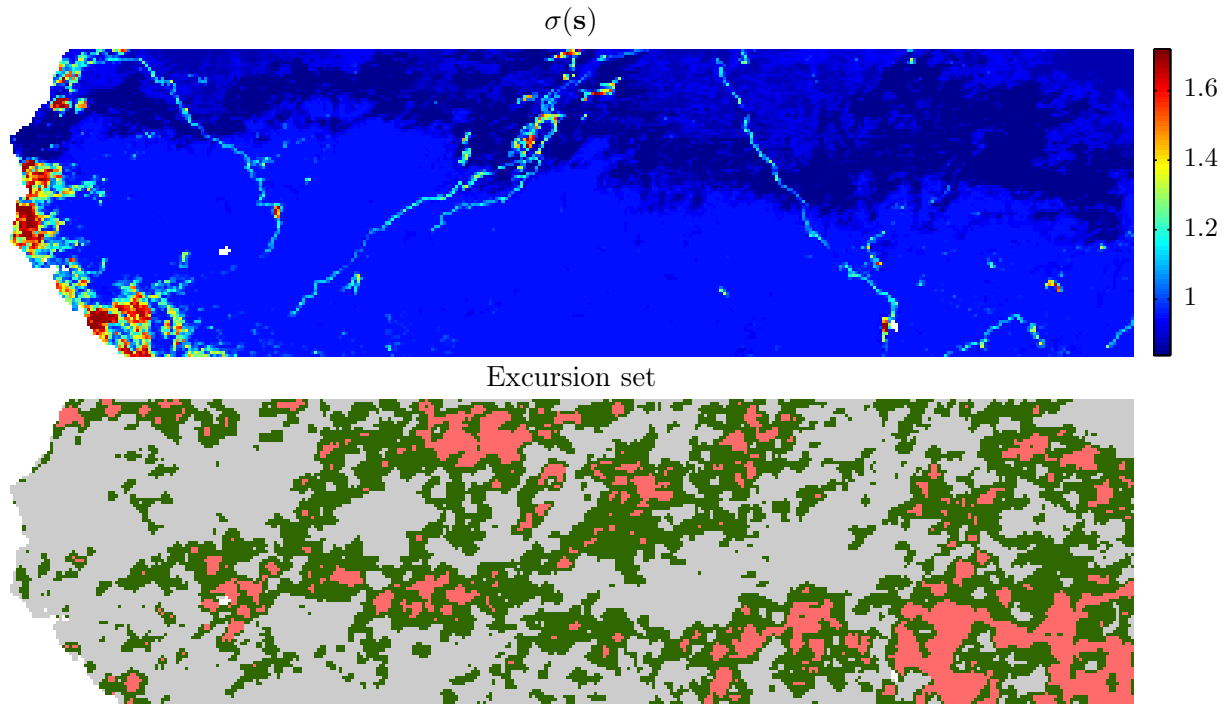

Figure 11: Results from the Sahel vegetation data. The top panel shows the estimated standard deviation of the measurement noise and the bottom panel shows the estimated excursion set $E_{0,0.05}^{+}(\mathbf{K}_2)$ in red and the point-wise positive significant trends in green.

(13). It is worth noting that these results look reasonable, with larger measurement errors in the coastal region and where there are forests and wetlands, and smaller measurement errors in desert and semi desert regions. Finally the bottom panel of Figure 11 shows the estimated excursion set $E_{0,0.05}^{+}(\mathbf{K}_2)$ in red and the point-wise positive significant trends in green. The interpretation of the result is that one with high certainty can conclude that the areas indicated in red have experienced an increase in vegetation over the studied time period. Hence, conclusions drawn by Eklundh and Olsson [2003] seem valid, also when taking the spatial dependency of the vegetation into account and when estimating the excursion sets controlling the family-wise error.

6 Discussion

Estimating excursion sets and uncertainty regions for contour curves for stochastic fields are difficult problems, both because of computational issues but also because it might not be clear how such uncertainty regions should be defined. In this work, we have given precise definitions for these uncertainty regions, introduced the concept of excursion functions as a visual tool for illustrating the uncertainty in these regions, and presented a method for calculating these quantities for latent Gaussian models.

The main idea behind the computational method is to use a parametric family for the excursion sets in combination with a sequential integration method to reduce the computational effort required when estimating the sets in practice. Tests on simulated data showed that the method is accurate, and two applications were presented to show that the method is applicable even to large environmental problems.

There are a number of extensions that could be made to this work. First of all, using the one-parameter family for the excursion sets gives a method that falls into the broad category of p-value thresholding methods for estimating simultaneous excursion sets. As previously mentioned, the important advantage with the method proposed here compared with other commonly used thresholding methods is that the correct joint distribution is used when selecting the threshold. The disadvantage is that the method is computationally more expensive than many of the standard thresholding methods. It would, therefore, be interesting to do a comparison with other similar methods with respect to the accuracy and computational complexity. Another interesting comparison would be to compare the uncertainty regions for contour curves produced by these methods to those of Lindgren and Rychlik [1995]. One could potentially also combine these methods with the work by Polfeldt [1999] to make statements on the quality of contour maps.

We also presented other parametric families that can be used to obtain more complicated methods for estimating the excursion sets, with the possibility of finding more precise estimates under the cost of higher computational complexity. Initial comparisons showed that there is not much gain in using these more complicated methods, but so far these comparisons have only been made using fairly simple latent models, and the gain is likely higher when the latent models are more complex. Hence, more studies are required to verify if this is the case and to investigate in what situations it is appropriate to use the simple one-parameter families. One possible advantage with the more complicated parametric families is when one has prior knowledge regarding the shape of the excursion sets. For example, if one knows that the excursion sets should be large contiguous regions, such knowledge could be incorporated using a two-parameter smoothing family.

As a final note, the method introduced here is available as a C-package with interfaces to both R and Matlab, see the supplementary material for the online version of the article for details.

Acknowledgements

with National Oceanic and Atmospheric Administration. The data were provided by the Earth Observing System Data and Information System (EOSDIS), Distributed Active Archive Center at Goddard Space Flight Center which archives, manages, and distributes this data set. The data used in the PM_{10} study was provided by the information system Aria Web Regione Piemonte and Arpa Piemonte. The authors are grateful to Johan Lindström and Daniel Simpson for valuable discussions on the subject of excursions and contour curve uncertainty sets, and to Peter Guttorp for highlighting the need for a thorough treatment of the subject.

A   Notes on the MCMC algorithm used in Example 3

To generate samples from the posterior distribution $\pi(\mathbf{x}|\mathbf{y})$ in Example 3, an MCMC algorithm is used. The algorithm is a random-walk Metropolis Hastings algorithm [Metropolis et al., 1953, Hastings, 1970] with proposal kernel

$$q(\{\mathbf{x}_{\text{old}}, \boldsymbol{\theta}_{\text{old}}\}, \{\mathbf{x}, \boldsymbol{\theta}\}) = \pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) q_\theta(\boldsymbol{\theta}_{\text{old}}, \theta).$$

A new proposal of the parameters $\boldsymbol{\theta} = (\log(\sigma), \log(\kappa), \log(\phi))$ is proposed based on the old value $\boldsymbol{\theta}_{\text{old}}$ using $\boldsymbol{\theta} \sim \mathsf{N}(\boldsymbol{\theta}_{\text{old}}, \boldsymbol{\Sigma}_\theta)$. Here $\boldsymbol{\Sigma}_\theta$ is a scaled version of the Hessian matrix evaluated at the maximum posterior estimate of $\boldsymbol{\theta}$. The scaling is selected as suggested by Gelman et al. [1996]. A new value for $\mathbf{x}$ is then proposed using the marginal posterior distribution $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ given by $\mathbf{x}|\mathbf{y}, \boldsymbol{\theta} \sim \mathsf{N}(\frac{1}{\sigma^2}\hat{\mathbf{Q}}^{-1}\mathbf{A}^\top\mathbf{y}, \hat{\mathbf{Q}})$. Here $\hat{\mathbf{Q}} = \mathbf{Q} + \frac{1}{\sigma^2}\mathbf{A}^\top\mathbf{A}$, where $\mathbf{Q}$ is the precision matrix for $\mathbf{x}$ and $\mathbf{A}$ is an observation matrix determined by the measurement locations. The acceptance probability simplifies to

$$\alpha_{\text{MCMC}} = \min\left(1, \frac{\pi(\boldsymbol{\theta}_{\text{old}}|\mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})}\right),$$

where the posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$ is given by

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{|\mathbf{Q}|^{\frac{1}{2}}\pi(\boldsymbol{\theta})}{|\hat{\mathbf{Q}}|^{\frac{1}{2}}|\sigma\mathbf{I}|}\exp\left(\frac{1}{2\sigma^2}\mathbf{y}^\top\left(\frac{\mathbf{A}\hat{\mathbf{Q}}^{-1}\mathbf{A}^\top}{\sigma^2} - \mathbf{I}\right)\mathbf{y}\right). \tag{14}$$

Since the proposal for $\mathbf{x}$ does not affect the acceptance probability, a new proposal for $\mathbf{x}$ is only generated if $\boldsymbol{\theta}$ is accepted. With only three parameters in the model, we achieve good mixing this way, but being an MCMC-procedure it is still highly computationally demanding since the calculation of the acceptance probability requires a few Cholesky factorizations and back substitutions based on the posterior precision matrix for $\mathbf{x}$.

References

R. J. Adler. *The Geometry of Random Fields.* Wiley, New York, 1981.

P.A. Agbu and M.E. James. *The NOAA/NASA Pathfinder AVHRR Land Data Set User's Manual.* Goddard Distributed Active Archive Center, NASA, Goddard Space Flight Center, Greenbelt, 1994.

M. M. Beaky, R. J. Scherrer, and J. V. Villumsen. Topology of large-scale structure in seeded hot dark matter models. *Astrophys. J.*, 387:443–448, 1992.

D. Bolin, J. Lindström, L. Eklundh, and F. Lindgren. Fast estimation of spatially dependent temporal vegetation trends using Gaussian Markov random fields. *Comput. Statist. and Data Anal.*, 53:2885–2896, 2009.

M. Cameletti, F. Lindgren, D. Simpson, and H. Rue. Spatio-temporal modeling of particulate matter concentration through the SPDE approach (submitted), 2012.

Michela Cameletti, Rosaria Ignaccolo, and Stefano Bande. Comparing spatio-temporal models for particulate matter in piemonte. *Environmetrics*, 22(8):985–996, 2011. ISSN 1099-095X. doi: 10. 1002/env.1139. URL `http://dx.doi.org/10.1002/env.1139`.

Martin A. Cohen, Sara D. Adar, Ryan W. Allen, Edward Avol, Cynthia L. Curl, Timothy Gould, David Hardie, Anne Ho, Patrick Kinney, Timothy V. Larson, Paul Sampson, Lianne Sheppard, Karen D. Stukovsky, Susan S. Swan, L-J Sally Liu, and Joel D. Kaufman. Approach to estimating participant pollutant exposures in the Multi-Ethnic Study of Atherosclerosis and air pollution (MESA air). *Environ. Sci. Technol.*, 43(13):4687–4693, 2009.

L. Eklundh and L. Olsson. Vegetation index trends for the African Sahel 1982–1999. *J. Geophys. Res.*, 30:1430–1433, 2003.

S. Finardi, R. De Maria, A. D'Allura, C. Cascone, G. Calori, and F. Lollobrigida. A deterministic air quality forecasting system for Torino urban area, Italy. *Environmental Modelling and Software*, 23 (3):344–355, 2008.

R. Furrer, R. Knutti, S. R. Sain, D. Nychka, and Meehl G. A. Spatial patterns of probabilistic temperature change projections from a multivariate bayesian analysis. *Geophys. Res. Lett.*, 34, 2007.

A Gelman, G O Roberts, and W R Gilks. Efficient Metropolis jumping rules. *Bayesian Statist.*, 5: 599–607, 1996.

A. Genz. Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.*, 1 (2):pp. 141–149, 1992.

A. Genz and F. Bretz. Comparison of methods for the computation of multivariate t probabilities. *J. Comput. Graph. Statist.*, 11(4):pp. 950–971, 2002.

A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*, volume 195 of *Lecture Notes in Statistics*. Springer, 2009.

A. Genz and D. Kahaner. The numerical evaluation of certain multivariate normal integrals. *J. Comput. Appl. Math.*, 16:255–258, 1986.

John Geweke. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities, 1991.

G. J. Gibson, C.A. Glasbey, and Elston D.A. Monte Carlo evaluation of multivariate normal integrals and sensitivity to variate ordering. In I.T. Dimov, B. Sendov, and P.S. Vassilevski, editors, *Advances in Numerical Methods and Aplications*, pages 120–126. World Scientific Publishing, River Edge, 1994.

V.A. Hajivassiliou. Simulation estimation methods for limited dependent variable models. Cowles Foundation Discussion Papers 1007, Cowles Foundation for Research in Economics, Yale University, December 1991. URL `http://ideas.repec.org/p/cwl/cwldpp/1007.html`.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

S. Holm. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6(2):pp. 65–70, 1979. ISSN 03036898. URL `http://www.jstor.org/stable/4615733`.

M.E. James and S.N.V. Kalluri. The Pathfinder AVHRR Land data set: An improved coarse resolution data set for terrestrial monitoring. *Internat. J. Remote Sensing*, 15:3347–3363, 1994.

M.P. Keane. 20 simulation estimation for panel data models with limited dependent variables. In *Econometrics*, volume 11 of *Handbook of Statistics*, pages 545 – 571. Elsevier, 1993. doi: 10.1016/S0169-7161(05)80055-0. URL `http://www.sciencedirect.com/science/article/pii/S0169716105800550`.

P.J. Lamb. Persistence of Subsaharan drought. *Nature*, 299:46–47, 1982.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *J. Roy. Statist. Soc. Ser. B Stat. Methodol.*, 73:423–498, 2011.

G. Lindgren and I. Rychlik. How reliable are contour curves? confidence sets for level contours. *Bernoulli*, 4(1):301–319, 1995.

J. Marchini and A. Presanis. Comparing methods of analyzing fMRI statistical parametric maps. *NeuroImage*, 22:1203–1213, 2003.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

S.E. Nicholson. Land surface process and Sahel climate. *Rev. of Geophys.*, 38:117–140, 2000.

Lennart Olsson. On the causes of famine – drought, desertification and market failure in the Sudan. *Ambio*, 22:395–403, 1993.

Thomas Polfeldt. On the quality of contour maps. *Environmetrics*, 10:785–790, 1999.

H. Rue and L. Held. *Gaussian Markov Random Fields; Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 2005.

H Rue, S Martino, and N Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J. Roy. Statist. Soc. Ser. B Stat. Methodol.*, 71(2):319–392, 2009.

S. R. Sain, R. Furrer, and N. Cressie. A spatial analysis of multivariate output from regional climate models. *Ann. Appl. Statist.*, 5:150–175, 2011.

Mark J. Schervish. Algorithm as 195: Multivariate normal probabilities with error bound. *J. Roy. Statist. Soc. Ser. C Appl. Statist.*, 33(1):pp. 81–94, 1984.